Software Requirements Specification (SRS)

Purpose

For the system to be able to restore datasets in case of a mishap causing data loss or corruption, we need to backup each dataset, and do it frequently to accommodate updates. Backups would also help in recovering earlier versions of a dataset if some unintended change was made.

Scope

This backup subsystem is supposed to be a part of a bigger project of managing datasets (involving development by many teams). The scope would therefore be limited to their requirements, and the individual components would have to be modified a little bit if someone else wants to use it as a general purpose backup system.

Assumptions

- If the dataset backup is done on a machine different from the dataset storage (which is ideal to prevent corruption), we assume that the API calls transferring data between servers would either be provided to us, or the configuration would be made known so we can plan accordingly.
- If backup is done on a timed basis, having 2 or more changes to the dataset in the time-interval would only preserve the last state of the dataset (and the state before the interval) as a backup version, while all changes in the middle would be lost. We assume that loss due to very frequent change is allowed.

Requirements

Functional requirements

- Backup and recover datasets stored on the system.
- Backup frequently to keep the datasets updated.
- Maintain version history to access previous versions of the datasets.
- Store only the modifications to datasets (between successive versions) to keep the size small.
- Be able to handle all file formats and directory structures.

- Easy and intuitive backup and recovery API calls should be available.
- Recovery should be available till any required date & time.

Non-functional requirements

- The backups should be reliable.
- The modifications should be consistently stored.
- Recovery should be fast.
- Backups should be frequent.
- Code should be modular.
- Should be able to handle multiple types of datasets.
- Data file type, directory structure, and format integrity should be maintained in each backup version.